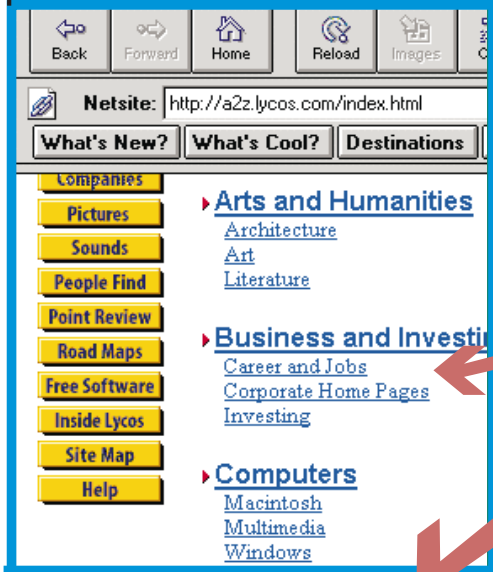
1998 WWW SEARCH THING

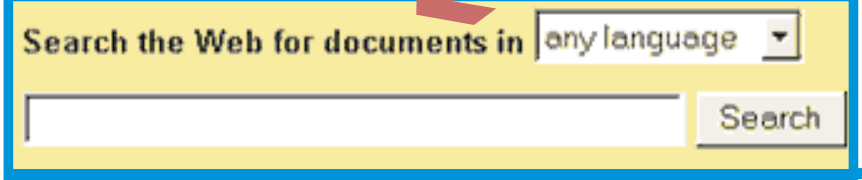
June 1998



Types of Search Sites

Directories for browsing.

Search Engines for pinpointing



Selecting Search Engines

- Try out several.
- Consider factors related to an engine's collecting function and its search interface.
- Use online help to get to know search features.

WWW Search Thing Authors

Roxanne Byrd
Brenda Epperson
Nancy Ralston
Erin Bauer
Ronnie Mevorach

WWW Search Thing Comments:
nralston@netserv.unmc.edu
800-338-7657 or 402-559-4326

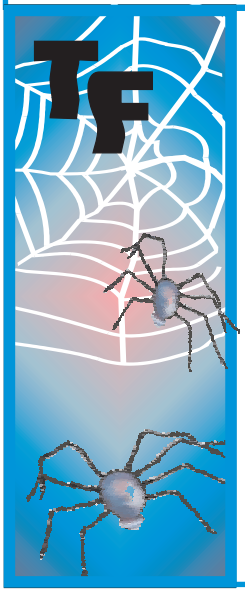
NN/LM
National Network of Libraries of Medicine
Midcontinental Region

Search Site Features To Look For

SIZE	Varies significantly from one engine to another. Large engines index from 30 - 100 million web pages; small engines index around 2 million pages.
FULL-TEXT SEARCH	Engines may index every word on a web page, just an abstract, or a condensed copy of a page. A full-text search engine indexes every word on pages.
BOOLEAN	You may be able to enlarge or restrict searches using operators like OR, AND and NOT. OR increases retrieval while AND and NOT reduce retrieval. Some engines allow you to select these operators or equivalent phrases like *All the words* from a menu. Others allow you to enter operators along with terms, and still others insert a (default) operator between words.
PROXIMITY	Frequently, you can increase precision by specifying how words should be positioned. For example, you may be able to indicate that two words be side-by-side or within 5 or 10 words of each other. Quotes and/or the NEAR operator are often used by searchers to specify adjacency/proximity. See entry samples included in the Search Thing grid.
TRUNCATION	Truncation usually refers to retrieving several terms by entering only a word stem followed by a symbol *(i.e., neoplas* to retrieve neoplasm, neoplasmic, or neoplastic)*. Truncation can be automatic, non-existent, or available only at the end of a word. Truncation symbols vary.
CASE SENSITIVITY	Case sensitive searches establish retrieval based on exact match of upper and lower case letters. Frequently, you can run case-insensitive searches by using NO-CAPS and case-sensitive searches by including capital letters.

Continued on page 4.

Search Features continued	
FIELD SEARCHING	The information about a search site that is found in a search engine's database exists as parts or fields such as the title, url, links, text, etc. Sometimes limitations may be made to these fields.
NESTING	When more than one boolean or proximity operator is used, an engine may process the entry according to a certain "batting order." Nesting allows you to group similar terms and specify the operation to be processed first. Parentheses are frequently used to indicate which part of the search should be done first.
ADVANCED FORM	Additional or different search features may be available via advanced form - only a click away.
LIMITING	Search sites such as Infoseek and Yahoo let users run one search of the entire database, then another search focusing on only documents found by a previous search.
WEIGHTED SEARCH	Some search engines allow users to control the relevance score for search results by preceding words with + or - signs to increase/decrease the importance of words entered in the search form.



Q: **True or False?** If you select the largest search engine, you can be sure that you've conducted a comprehensive search of the World Wide Web.

A: **FALSE.** No one search site indexes every Web page.

Q: **True or False?** A search engine finds information for its database by accepting listings sent by authors or by getting the information from automated robots sometimes called "Web crawlers" or "spiders".

A: **TRUE**

Q: **True or False?** When a computerized web crawler roams the Internet looking for information to add to an engine's database, it sends back just the title and URL of each page it visits.

A: **FALSE.** A crawler can send back the above, the entire text of each page, or data contained in document fields.

Search Site Components

- The **database -** where records about web pages and Internet resources reside.
- The **automated robot or indexing system** - used to gather information about sites/resources.
- The **search interface -** which allows users to tap the resources of the database *(includes search form, available search features and presentation of results).*

Presentation of Results

The way in which a search engine ranks and displays search results is not the primary focus of this document, but it is an issue well-worth examination. Many search engines return so many items that the typical user can be discouraged when it becomes necessary to sift through several results pages. Often there is an elaborate scheme established to rank items retrieved. Search engines vary regarding these algorithms, as well as the helpfulness of summaries for items retrieved.

This handout has been designed to introduce medical/health professionals and students to World Wide Web search sites and techniques. Information has been selectively chosen. To bring search site changes to the attention of this document's authors, or to request copies, contact Nancy Ralston at the National Network of Libraries of Medicine - Midcontinental Region. E-mail: nralston@netserv.unmc.edu Phone: 800-338-7657